The Annals of Probability
1996, Vol. 24, No. 4, 2128–2140

NEW DONSKER CLASSES[1]

BY AAD VAN DER VAART

Vrije Universiteit

Several classes of functions are shown to be Donsker by an argument based on partitioning the sample space. One example is the class of all nondecreasing functions $f\colon \mathbb{R} \to \mathbb{R}$ such that $0 \le f \le F$ for a given function F with $\int F^2\,dP/\sqrt{1-P} < \infty$.

1. A general result. Let P be a probability measure on the measurable space $(\mathscr{X}, \mathscr{A})$ and $\mathscr{F}$ a class of measurable functions $f\colon \mathscr{X} \to \mathbb{R}$ such that $Pf^2 < \infty$ for every $f \in \mathscr{F}$. The class $\mathscr{F}$ is called P-Donsker if the empirical process $\{\mathbb{G}_n f\colon f \in \mathscr{F}\}$ converges in distribution to a tight Brownian bridge process $\mathbb{G}$ in the space $l^\infty(\mathscr{F})$ of uniformly bounded functions on $\mathscr{F}$. Here $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ with $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ denoting the empirical measure of an i.i.d. sample from P. Convergence in distribution is understood in the sense that $\mathrm{E}^* h(\mathbb{G}_n) \to \mathrm{E} h(\mathbb{G})$ for every continuous bounded function $h\colon l^\infty(\mathscr{F}) \to \mathbb{R}$; compare Dudley (1985).

Let $\mathscr{X} = \bigcup_{j=1}^\infty \mathscr{X}_j$ be a partition of $\mathscr{X}$ into measurable sets and let $\mathscr{F}_j$ be the class of functions $f 1_{\mathscr{X}_j}$ when f ranges over $\mathscr{F}$. If the class $\mathscr{F}$ is Donsker, then each class $\mathscr{F}_j$ is Donsker. This is not obvious from the definition of a Donsker class, but can be proved by extending Corollary 14.8 of Giné and Zinn (1986a) with the measurability conditions taken care of along the lines of Talagrand (1987), because a restriction is a contraction in the sense that $|(f 1_{\mathscr{X}_j})(x) - (g 1_{\mathscr{X}_j})(x)| \le |f(x) - g(x)|$ for every x. We shall not use this claim in this note, but we are interested in a converse of this statement. While the sum of infinitely many Donsker classes need not be Donsker, it is clear that if each $\mathscr{F}_j$ is Donsker and the classes $\mathscr{F}_j$ become suitably small as $j \to \infty$, then $\mathscr{F}$ is Donsker. The following precise version of this principle enables us to deal with a number of interesting examples. Let $\|G\|_{\mathscr{F}} = \sup\{|Gf|\colon f \in \mathscr{F}\}$ be the norm of the process G. Let F be an envelope function of the class $\mathscr{F}$, a function such that $|f| \le F$ for every f.

THEOREM 1.1. *For each j let the class of functions $\mathscr{F}_j$ be Donsker and suppose that*

$$\mathrm{E}_P^* \|\mathbb{G}_n\|_{\mathscr{F}_j} \le C c_j$$

for a constant C not depending on j or n. If $\sum_{j=1}^\infty c_j < \infty$ and $P^* F < \infty$, then the class $\mathscr{F}$ is P-Donsker.

Received November 1994; revised November 1995.
[1]Partially supported by NATO CRG 920614.
AMS 1991 *subject classification*. 60F17.
Key words and phrases. Bracketing number, covering number, entropy, Donsker class, empirical central limit theorem.

PROOF. We can assume without loss of generality that the class $\mathscr{F}$ contains the constant function 1.

The assumption that $\mathscr{F}_j$ is Donsker entails that the sequence of empirical processes indexed by $\mathscr{F}_j$ converges in distribution in $l^\infty(\mathscr{F}_j)$ to a tight Brownian bridge $\mathbb{H}_j$ for each j. This implies that $\mathrm{E}^*\|\mathbb{G}_n\|_{\mathscr{F}_j} \to \mathrm{E}\|\mathbb{H}_j\|_{\mathscr{F}_j}$; compare Giné and Zinn (1986a). Thus $\mathrm{E}\|\mathbb{H}_j\|_{\mathscr{F}_j} \le Cc_j$. Let Z_j be a standard normal variable independent of $\mathbb{H}_j$ constructed on the same probability space. Since $\sup_{f \in \mathscr{F}_i}|Pf| < \infty$ the process $f \to \mathbb{W}_j(f) = \mathbb{H}_j(f) + Z_j Pf$ is well defined on $\mathscr{F}_j$ and takes its values in $l^\infty(\mathscr{F}_j)$. We can construct these processes for different j as independent random elements on a single probability space. Then the series $\mathbb{W}(f) = \sum_{j=1}^\infty \mathbb{W}_j(f1_{\mathscr{X}_j})$ converges in second mean for every f and satisfies $\mathrm{E}\mathbb{W}(f)\mathbb{W}(g) = Pfg$ for every f and g. Thus the series defines a version of a Brownian motion process. Since each of the processes $\{\mathbb{W}_j(f1_{\mathscr{X}_j})\colon f \in \mathscr{F}\}$ has bounded and uniformly continuous sample paths with respect to the $L_2(P)$-seminorm, so have the partial sums $\mathbb{W}_{\le k} = \{\sum_{j=1}^k \mathbb{W}_j(f1_{\mathscr{X}_j})\colon f \in \mathscr{F}\}$. Furthermore,

$$\begin{aligned}\mathrm{E}\sup_{f\in\mathscr{F}}\left|\sum_{j>k}\mathbb{W}_j(f1_{\mathscr{X}_j})\right| &\le \sum_{j>k}\left(\mathrm{E}\|\mathbb{H}_j\|_{\mathscr{F}_j} + \mathrm{E}|Z_j|P^*F1_{\mathscr{X}_j}\right)\\ &\le C\sum_{j>k}c_j + \sqrt{2/\pi}\,P^*F1_{\cup_{j>k}\mathscr{X}_j}.\end{aligned}$$

This converges to zero as $k \to \infty$. Thus the series $\mathbb{W}(f) = \sum_{j=1}^\infty \mathbb{W}_j(f1_{\mathscr{X}_j})$ converges in mean in the space $l^\infty(\mathscr{F})$. By the Itô–Nisio theorem [e.g., Ledoux and Talagrand (1991), Theorem 2.4] it also converges almost surely. We conclude that almost all sample paths of the process $\mathbb{W}$ are bounded and uniformly continuous. Since $\mathrm{E}\|\mathbb{W}\|_{\mathscr{F}} < \infty$, the class $\mathscr{F}$ is totally bounded in $L_2(P)$ by Sudakov's inequality [e.g., Ledoux and Talagrand (1991), Theorem 3.18]. Hence $\mathbb{W}$ is a tight version of a Brownian motion process indexed by $\mathscr{F}$. The process $\mathbb{G}(f) = \mathbb{W}(f) - \mathbb{W}(1)Pf$ defines a tight Brownian bridge process indexed by $\mathscr{F}$. We have proved that $\mathscr{F}$ is pre-Gaussian.

For each k set $\mathbb{G}_{n,\le k}(f) = \mathbb{G}_n(f1_{\cup_{j\le k}\mathscr{X}_j})$. The continuity modulus of the process $\{\mathbb{G}_{n,\le k}(f);\ f \in \mathscr{F}\}$ is bounded by the continuity modulus of the empirical process indexed by the sum class $\sum_{j\le k}\mathscr{F}_j$. [The class of all functions $x \to f_1(x) + \cdots + f_k(x)$ as f_i ranges over $\mathscr{F}_i$ for $i = 1,\dots,k$.] By Proposition 2.6 of Alexander (1987) the sum of finitely many Donsker classes is Donsker. (Since Alexander does not give the proof in his paper, his proposition is restated and proved in the Appendix.) We can conclude that the sequence $(\mathbb{G}_{n,\le k})_{n=1}^\infty$ is asymptotically tight in $l^\infty(\mathscr{F})$. Considering the marginal distributions, we can conclude that the sequence converges in distribution to

$$\mathbb{G}_{\le k}(f) = \mathbb{W}_{\le k}(f) - d_k\mathbb{W}_{\le k}(1)Pf1_{\cup_{j\le k}\mathscr{X}_j}$$

as $n \to \infty$ for each fixed k, for d_k a solution to the equation $d^2P(\cup_{j\le k}\mathscr{X}_j) - 2d = -1$.

As $k \to \infty$ the sequence $\mathbb{G}_{\le k}$ converges almost surely, whence in distribution to $\mathbb{G}$ in $l^\infty(\mathscr{F})$. Weak convergence to a tight limit is metrizable by (for instance) the dual bounded Lipschitz metric; compare Theorem B of Dudley (1990). Since the array $\mathbb{G}_{n,\le k}$ converges along every row to limits that converge to $\mathbb{G}$, there exist integers $k_n \to \infty$ such that the sequence $\mathbb{G}_{n,\le k_n}$ converges in distribution to $\mathbb{G}$. We also have that the sequence $\mathbb{G}_n - \mathbb{G}_{n,\le k_n}$ converges in outer probability to zero, since

$$\mathrm{E}^*\|\mathbb{G}_n - \mathbb{G}_{n,\le k_n}\|_{\mathscr{F}} \le C \sum_{j>k_n} c_j \to 0.$$

An application of Slutsky's lemma completes the proof. □

An upper bound on the mean of the maximum of the empirical process may be based on uniform entropy numbers or bracketing entropy numbers. Let F be a measurable envelope function of the class $\mathscr{F}$. Let $\|f\|_{P,2}$ denote the $L_2(P)$-norm of a function f.

Given a pair of functions $l \le u$, the bracket $[l, u]$ consists of all functions f with $l \le f \le u$. The bracketing number $N_{[\,]}(\varepsilon, \mathscr{F}, L_2(P))$ is the minimal number of brackets $[l, u]$ of size $P(u-l)^2$ smaller than ε^2 needed to cover $\mathscr{F}$. There exists a universal constant C such that for any class of functions $\mathscr{F}$

$$\mathrm{E}_P^*\|\mathbb{G}_n\|_{\mathscr{F}} \le C\int_0^1 \sqrt{1 + \log N_{[\,]}(\varepsilon\|F\|_{P,2}, \mathscr{F}, L_2(P))}\, d\varepsilon \|F\|_{P,2}. \tag{1}$$

This is proved by Pollard (1989) in an unpublished manuscript. It follows from recasting the maximal inequalities for tail probabilities of Ossiander (1985) and Arcones and Giné (1993) into inequalities for first moments. For completeness we have included a proof in Section 6.

The covering number $N(\varepsilon, \mathscr{F}, L_2(P))$ is the minimal number of $L_2(P)$ balls of size ε needed to cover $\mathscr{F}$. Under measurability conditions

$$\mathrm{E}_P^*\|\mathbb{G}_n\|_{\mathscr{F}} \le C\int_0^1 \sup_Q \sqrt{1 + \log N(\varepsilon\|F\|_{Q,2}, \mathscr{F}, L_2(Q))}\, d\varepsilon \|F\|_{P,2}, \tag{2}$$

where the supremum is taken over all finitely discrete probability measures Q. See Kim and Pollard (1990). This bound is valid only under some measurability conditions on the class $\mathscr{F}$. It suffices that the observations $X_1, X_2, \ldots$ are defined as the coordinate projections on the product space $(X^\infty, \mathscr{A}^\infty)$ and that the map

$$(X_1, \ldots, X_n) \to \left\| \sum_{i=1}^n e_i f(X_i) \right\|_{\mathscr{F}}$$

is measurable for every $(e_1, \ldots, e_n) \in \{0, 1\}^n$ and every n. The examples in this note concern separable classes of functions, for which no measurability difficulties arise.

For many classes of functions $\mathscr{F}$ the integrals in the preceding upper bounds are finite. If they are uniformly finite for the classes $\mathscr{F}_j$, then the theorem applies with c_j equal to the norm of the envelope of $\mathscr{F}_j$. In that case the condition $\sum_{j=1}^\infty \|F_j\|_{P,2} < \infty$ implies that $\mathscr{F}$ is Donsker. Note that convergence of this series does not require that the envelopes F_j become small;

convergence to zero of the measure $P(\mathscr{X}_j)$ of the jth set in the partition may compensate for an increasing envelope.

2. Smooth functions. Classes of functions that are smooth up to order α are defined as follows. For any vector $k = (k_1, \dots, k_d)$ of d integers define the differential operator

$$D^k = \frac{\partial^{k.}}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}},$$

where $k. = \sum k_i$. For $0 < \alpha < \infty$ let $\lfloor \alpha \rfloor$ be the greatest integer strictly smaller than α. Then for a function $f: \mathscr{X} \subset \mathbb{R}^d \to \mathbb{R}$ let

$$\|f\|_\alpha = \max_{k. \le \lfloor \alpha \rfloor} \sup_x \left| D^k f(x) \right| + \max_{k. = \lfloor \alpha \rfloor} \sup_{x, y} \frac{\left| D^k f(x) - D^k f(y) \right|}{\|x - y\|^{\alpha - \lfloor \alpha \rfloor}},$$

where the suprema are taken over all x, y in the interior of $\mathscr{X}$ with $x \neq y$. Let $C_M^\alpha(\mathscr{X})$ be the set of all continuous functions $f: \mathscr{X} \to \mathbb{R}$ with $\|f\|_\alpha \le M$. Note that for $\alpha \le 1$ this class consists of bounded functions f that satisfy a Lipschitz condition.

Let $\mathbb{R}^d = \bigcup_{j=1}^\infty \mathscr{X}_j$ be a partition of $\mathbb{R}^d$ into uniformly bounded, convex sets with nonempty interior. Consider the class of $\mathscr{F}$ of functions such that the class $\mathscr{F}_j$ of restrictions is contained in $C_{M_j}^\alpha(\mathscr{X}_j)$ for each j for given constants M_j.

Kolmogorov and Tikhomirov (1961) computed the entropy of the classes of $C_M^\alpha(\mathscr{X})$ for the uniform norm. As a consequence of their results there exists a constant K depending only on α, d and the diameter of $\mathscr{X}$ such that for every measure P and every $\varepsilon > 0$,

$$\log N_{[\,]}\Big(\varepsilon M P(\mathscr{X})^{1/2}, C_M^\alpha(\mathscr{X}), L_2(P)\Big) \le K\left(\frac{1}{\varepsilon}\right)^{d/\alpha}.$$

For $\alpha > d/2$ the exponent of $(1/\varepsilon)$ is strictly less than 2. Thus for $\alpha > d/2$ each of the classes $\mathscr{F}_j$ is Donsker by the bracketing central limit theorem of Ossiander (1985). Moreover, in view of (1) or (2) there exists a constant C not depending on j or n such that

$$\mathrm{E}^* \|\mathbb{G}_n\|_{\mathscr{F}_i} \le C M_j P(\mathscr{X}_j)^{1/2}.$$

This yields the following corollary to Theorem 1.1.

COROLLARY 2.1. *If $\alpha > d/2$ and $\sum_{j=1}^\infty M_j P(\mathscr{X}_j)^{1/2} < \infty$, then the class $\mathscr{F}$ is P-Donsker.*

For dimension $d = 1$ and uniform bounds $M_j \equiv 1$ this result is obtained by Giné and Zinn (1986a, b) by a much more complicated proof. Giné and Zinn also show (in their case) that convergence of the series is necessary for the class $\mathscr{F}$ to be pre-Gaussian. This extends to the present situation. It may be noted that the required smoothness $\alpha > d/2$ is necessary already for $C_1^\alpha(\mathscr{X})$

to be Donsker for a single set $\mathscr{X} = \mathscr{X}_j$; compare Dudley [(1984), Theorem 8.1.1]. As regards the convergence of the series we have the following result, which is closely connected to the Borisov–Durst theorem; compare Dudley [(1984), Theorem 6.3.1]. Call a class of measurable functions $\mathscr{F}$ pre-Gaussian if it is contained in $\mathscr{L}_2(P)$ and there exists a tight version of the Brownian bridge indexed by $\mathscr{F}$.

LEMMA 2.2. *If the class* $\mathscr{F} = \{\sum_{j\in A} M_j 1_{\mathscr{X}_j}\colon A \subset \mathbb{N}\}$ *is P-pre-Gaussian, then* $\sum_{j=1}^{\infty} M_j P(\mathscr{X}_j)^{1/2} < \infty$.

PROOF. Since each $f \in \mathscr{F}$ is square integrable by assumption, it follows that $\sum_{j=1}^{\infty} M_j^2 P(\mathscr{X}_j) = P(\sum M_j 1_{\mathscr{X}_j})^2 < \infty$. This implies that $\sup_{f\in\mathscr{F}} P|f| < \infty$ and we can conclude that $\mathscr{F}$ can index not only a tight Brownian bridge, but also a tight Brownian motion $\mathbb{Z}$. It is well known that this can be represented as a series

$$\mathbb{Z} = \sum_{j=1}^{\infty} (Pf\psi_j) Z_j \quad \text{a.s.}$$

for an i.i.d. sequence $Z_1, Z_2, \dots$ of standard normal variables and any orthonormal set $\psi_1, \psi_2, \dots,$ in $L_2(P)$ whose closed linear span contains $\mathscr{F}$; compare Dudley [(1985), Theorem 5.1 and its proof]. The series converges uniformly in $f \in \mathscr{F}$. Thus

$$\sup_f \sum_{j=1}^{\infty} (Pf\psi_j) Z_j < \infty \quad \text{a.s.}$$

Apply this with the functions $\psi_j = P(\mathscr{X}_j)^{-1/2} 1_{\mathscr{X}_j}$ to find that

$$\sum_{j=1}^{\infty} M_j P(\mathscr{X}_j)^{1/2} Z_j^+ = \sup_{a\subset\mathbb{N}} \sum_{j\in A} M_j P(\mathscr{X}_j)^{1/2} Z_j < \infty \quad \text{a.s.}$$

Similar reasoning gives the same statement concerning the negative parts of the Z_j. Thus the series $\sum_{j=1}^{\infty} M_j P(\mathscr{X}_j)^{1/2}|Z_j|$ converges almost surely. By the three series theorem

$$\sum_{j=1}^{\infty} M_j P(\mathscr{X}_j)^{1/2} \mathrm{E}|Z_j| 1\Big\{M_j P(\mathscr{X}_j)^{1/2}|Z_j| \le 1\Big\} < \infty.$$

Since $M_j P(\mathscr{X}_j)^{1/2} \to 0$, the expectations are bounded away from zero and we obtain the desired result. □

3. Monotone functions. Consider the class $\mathscr{F}$ of all nondecreasing functions $f\colon \mathbb{R} \to \mathbb{R}$ such that $0 \le f \le F$, for a given nondecreasing function F. For $F \equiv M$ for a constant M, this class satisfies

$$\log N_{[\,]}(\varepsilon M, \mathscr{F}, L_2(P)) \le K\left(\frac{1}{\varepsilon}\right),$$

for a universal constant K and every probability measure P. [See Birman and Solomjak (1967).] (The upper bound $(1/\varepsilon)\log(1/\varepsilon)$ is elementary and suffices for our purposes.) Thus a uniformly bounded class of monotone functions is Donsker by the bracketing central limit theorem of Ossiander (1985). We shall extend this well known result to the case of unbounded $\mathscr{F}$. One application can be found in Gill (1994).

The restriction $\mathscr{F}_j$ of the class $\mathscr{F}$ to a compact interval $\mathscr{X}_j$ is contained in the class of monotone nondecreasing functions bounded by $F_j = \sup\{F(x): x \in \mathscr{X}_j\}$. Thus, given any partition $\mathbb{R} = \bigcup_{j=1}^{\infty}\mathscr{X}_j$ into intervals $\mathscr{X}_j$, the class $\mathscr{F}_j$ is Donsker. Moreover, by (1) or (2),

$$\mathrm{E}^*\|\mathbb{G}_n\|_{\mathscr{F}_j} \le C\|F_j\|_{P,2},$$

for a universal constant C. The theorem yields the following result.

Corollary 3.1. *If $\int F\,dG/\sqrt{1-G} < \infty$ for the cumulative distribution function G, then $\mathscr{F}$ is G-Donsker. In particular, it suffices that $\int F^{2+\delta}\,dG < \infty$ for some $\delta > 0$.*

Proof. The problem can be reduced to the case of uniform $[0,1]$ observations by the quantile transformation. If $G^{-1} = \inf\{x: G(x) \ge u\}$ is the quantile function of G, then the class $\mathscr{F}\circ G^{-1}$ is Donsker with respect to the uniform measure if and only if $\mathscr{F}$ is G-Donsker. The class $\mathscr{G} = \mathscr{F}\circ G^{-1}$ consists of monotone functions $g: [0,1] \to \mathbb{R}$ with $0 \le g \le F\circ G^{-1}$. Furthermore,

$$\int_0^1 \frac{F\circ G^{-1}}{\sqrt{1+u}}\,du \le \int \frac{F}{\sqrt{1-G}}\,dG.$$

Thus assume without loss of generality that G is the uniform measure.

Now use the theorem with the partition into the intervals $\mathscr{X}_j = (x_j, x_{j+1}]$ with $x_j = 1 - 2^{-j}$ for each integer $j \ge 0$. The condition of the theorem involves the series

$$\sum_{j=0}^{\infty}\|F_j\|_{P,2} = \sum_{j=1}^{\infty}\frac{F(1-2^{-j})}{\sqrt{1-(1-2^{-j})}}2^{-j} \le 2\int_0^1 \frac{F(u)}{\sqrt{1-u}}\,du.$$

The corollary follows by the quantile transformation. $\square$

By partial integration we have (for F right continuous)

$$\int \frac{F}{\sqrt{1-G}}\,dG = -2F\sqrt{1-G}\Big|_{-\infty}^{\infty} + 2\int\sqrt{1-G_-}\,dF$$
$$\le 2F(-\infty) + 2\int_0^1 G\big[F^{-1}(u),\infty\big)^{1/2}\,du.$$

This shows that the condition of the corollary is satisfied if the envelope F has a finite $L_{2,1}$-norm $\|F\|_{2,1} = \int_0^\infty G(y: F(y) \ge x)^{1/2}\,dx$. At the time of revision of this paper we have learned that in this form the corollary has also been obtained independently and by a different method by Dudley and Koltchinskii (1994).

Our proof uses only that the functions are monotone on each of the partitioning sets, not their global monotonicity, albeit that the partitioning is chosen to optimize the resulting moment condition on the envelope. Theorem 1.1 is also applicable (for instance) to the class of all functions that are piecewise monotone with changes of direction at a countable number of given points at most. The class $\mathscr{F}$ will be Donsker provided the envelope satisfies a moment condition, which will however depend on the points of change of direction. (It is not claimed that the moment condition would be optimal, as it is in the case of Corollary 3.1.)

4. Closed convex subsets of $\mathbb{R}^2$. Consider the class $\mathscr{C}$ of all closed convex subsets of $\mathbb{R}^2$. It is well known that the intersection of this class with any bounded, convex set is P-Donsker for every probability measure P with a bounded Lebesgue density p. The full class $\mathscr{C}$ is Donsker if P has a density with small tails. We shall derive this from Theorem 1.1 applied to the class $\mathscr{F}$ of all indicator functions of sets in $\mathscr{C}$.

For any square $\mathscr{X}$ there exists a constant $K_{\mathscr{X}}$ that depends on the Lebesgue measure of $\mathscr{X}$ only, such that

$$\log N_{[\,]}\big(\varepsilon, \mathscr{C} \cap \mathscr{X}, L_2(P)\big) \le K_{\mathscr{X}} \|p\|_{\mathscr{X}}^{1/2} \left(\frac{1}{\varepsilon}\right).$$

Here $\|p\|_{\mathscr{X}}$ is the supremum of p over $\mathscr{X}$. This can be derived from the entropy of this class with respect to the Hausdorff metric; compare Dudley (1984). Choose a partition $\mathbb{R}^2 = \bigcup_{j=1}^{\infty} \mathscr{X}_j$ into squares of a fixed size. Then the preceding upper bound combined with (1) yields the existence of a constant C not depending on j or n such that

$$\mathrm{E}^* \|\mathbb{G}_n\|_{\mathscr{F}_j} \le C \|p\|_{\mathscr{X}_j}^{1/2}.$$

Thus the theorem yields the following corollary.

COROLLARY 4.1. *Suppose $\mathbb{R}^2 = \bigcup_{j=1}^{\infty} \mathscr{X}_j$ is a partition into squares of fixed size. Then $\mathscr{C}$ is Donsker if $\sum \|p\|_{\mathscr{X}_j}^{1/2} < \infty$. In particular, this is the case if $(1 + |xy|^{2+\delta}) p(x, y)$ is bounded for some $\delta > 0$.*

5. Functions of bounded variation. Consider the class $\mathscr{F}$ of all functions $f\colon \mathbb{R}^d \to \mathbb{R}$ of the type $f(x) = \mu(-\infty, x]$ for a signed Borel measure μ such that $|\mu|(-\infty, x] \le F(x)$ for a fixed, measurable, nondecreasing function F. The class $\mathscr{F}_{a,b}$ of restrictions of functions in $\mathscr{F}$ to the interval $(a, b]$ are cumulative distribution functions of signed measures of total variation bounded by $F(b)$. By the Glivenko–Cantelli theorem these functions can be approximated uniformly by cumulative distribution functions of discrete measures of the type $x \to \sum_i p_i 1_{x \le t}$, where $\sum_i |p_i| \le 2F(b)$. Thus the class is in the uniformly closed symmetric convex hull of the cells $(-\infty, t]$. The collection of cells $(-\infty, t]$ is VC and hence has polynomial covering numbers. An extension of Theorem 5.1 of Dudley (1987) yields that

$$\log N\big(\varepsilon F(b) Q^{1/2}(a,b], \mathscr{F}_{a,b}, L_2(Q)\big) \le K\left(\frac{1}{\varepsilon}\right)^r$$

for every measure Q, for some $r < 2$ depending on d only and for a constant K that depends on r only. The uniform entropy central limit theorem of Pollard (1982) implies that $\mathscr{F}_{a,b}$ is Donsker. Furthermore, by (2) there exists a constant C depending on d only, such that

$$\mathrm{E}^*\|\mathbb{G}_n\|_{\mathscr{F}_{a,b}} \le CF(b) P^{1/2}(a,b].$$

The theorem yields the following corollary.

COROLLARY 5.1. *Suppose $\mathbb{R}^d = \bigcup_{j=1}^{\infty}(a_j, b_j]$ is an arbitrary partition into intervals. Then $\mathscr{F}$ is Donsker if $\sum_{j=1}^{\infty} F(b_j) P^{1/2}(a_j, b_j] < \infty$.*

APPENDIX

In this appendix we supply a proof of inequality (1) and a proof that the sum of two Donsker classes is Donsker. These results are not new, but published proofs are apparently not available.

Inequality (1) was first formulated and proved by Pollard (1989) in an unpublished paper. Our proof uses the notation for the chaining argument introduced by Arcones and Giné (1993), though we chain using means rather than exponential inequalities. We need the following lemma.

LEMMA A.1. *Let $X_1, \dots, X_m$ be arbitrary random variables that satisfy the tail bound*

$$P(|X_i| > x) \le 2\exp\left(-\frac{1}{2}\frac{x^2}{b + ax}\right)$$

for all x and i and fixed $a, b > 0$. Then

$$E \max_{1 \le i \le m} |X_i| \le K\Big(a \log(1+m) + \sqrt{b}\sqrt{\log(1+m)}\Big)$$

for a universal constant K.

PROOF. The condition implies the upper bound $2\exp(-x^2/(4b))$ on $\mathrm{P}(|X_i| > x)$ for every $x \le b/a$ and the upper bound $2\exp(-x/(4a))$ for all other positive x. Consequently, the same upper bounds hold for *all* $x > 0$ for the probabilities $\mathrm{P}(|X_i| 1\{|X_i| \le b/a\} > x)$ and $\mathrm{P}(|X_i| 1\{|X_i| > b/a\} > x)$, respectively. By partial integration we next conclude that for a sufficiently large universal constant K,

$$\mathrm{E}\psi_2\left(\frac{|X_i| 1\{|X_i| \le b/a\}}{K\sqrt{b}}\right) \le 1; \qquad \mathrm{E}\psi_1\left(\frac{|X_i| 1\{|X_i| > b/a\}}{Ka}\right) \le 1,$$

where $\psi_p = \exp x^p - 1$. Thus by the triangle inequality and inequality (2.10) in Arcones and Giné (1993),

$$\mathrm{E}\max_i |X_i| \le \psi_2^{-1}(m)K\sqrt{b} + \psi_1^{-1}(m)Ka.$$

This is the assertion of the lemma. $\square$

By Bernstein's inequality the empirical process $\mathbb{G}_n f$ satisfies the tail bound of the preceding lemma with $b = Pf^2$ and $a = (1/3)\|f\|_\infty/\sqrt{n}$. Thus we obtain that for any finite class $\mathscr{F}$ of measurable square integrable functions with $|\mathscr{F}|$ elements,

$$\mathrm{E}\|\mathbb{G}_n\|_{\mathscr{F}} \le K\left(\max_f \frac{\|f\|_\infty}{\sqrt{n}}\log(1+|\mathscr{F}|) + \max_f \|f\|_{P,2}\sqrt{\log(1+|F|)}\right).$$

THEOREM A.2. *Any class $\mathscr{F}$ of measurable functions with square integrable envelope function F satisfies* (1).

PROOF. We use the notation $\lesssim$ for "smaller than, up to a universal constant" and use * to denote outer expectation and minimal measurable cover functions, as defined by Dudley (1985).

Fix an integer q_0 such that $2^{-q_0} < \|F\|_{P,2} \le 2^{-q_0+1}$. For every integer $q \ge q_0$, construct a nested sequence of partitions $\mathscr{F} = \bigcup_{i=1}^{N_q}\mathscr{F}_{qi}$ such that

$$P^* \sup_{f,g\in\mathscr{F}_{qi}} |f-g|^2 < 2^{-2q} \quad \text{for every } i.$$

This may be achieved by first selecting for each q a minimal number of brackets of size 2^{-q} that cover $\mathscr{F}$, disjointifying the resulting subsets and finally intersecting for each q all partitions of levels $q_0, q_0+1, \ldots, q$ to obtain a *nested* sequence of partitions. Thus the number $N_q - 1$ of subsets in the qth partition can be chosen to satisfy

$$\log N_q \le \sum_{r=q_0}^{q} \log\bigl(1 + N_{[\,]}(2^{-r}, \mathscr{F}, \|\cdot\|_{P,2})\bigr). \tag{3}$$

Choose for each q a fixed element f_{qi} from each partitioning set $\mathscr{F}_{qi}$ and set

$$\pi_q f = f_{qi}, \quad \text{if } f \in \mathscr{F}_{qi},$$
$$\Delta_q f = \Bigl(\sup_{f,g\in\mathscr{F}_{qi}} |f-q|\Bigr)^*, \quad \text{if } f\in\mathscr{F}_{qi}.$$

Note that $\pi_q f$ and $\Delta_q f$ run through a set of N_q functions if f runs through $\mathscr{F}$.

Define for each fixed n and $q \geq q_0$ numbers and indicator functions

$$
\begin{aligned}
a_q &= 2^{-q}/\sqrt{\log N_{q+1}}, \\
A_{q-1} f &= 1\left\{\Delta_{q_0} f \leq \sqrt{n}\, a_{q_0}, \ldots, \Delta_{q-1} f \leq \sqrt{n}\, a_{q-1}\right\}, \\
B_q f &= 1\left\{\Delta_{q_0} f \leq \sqrt{n}\, a_{q_0}, \ldots, \Delta_{q-1} f \leq \sqrt{n}\, a_{q-1}, \Delta_q f > \sqrt{n}\, a_q\right\}, \\
B_{q_0} f &= 1\left\{\Delta_{q_0} f > \sqrt{n}\, a_{q_0}\right\}.
\end{aligned}
$$

Note that $A_q f$ and $B_q f$ are constant in f on each of the partitioning sets $\mathscr{F}_{qi}$ at level q, because the partitions are nested. Now decompose, pointwise in x (which is suppressed in the notation):

$$
\begin{aligned}
f - \pi_{q_0} f = (f - \pi_{q_0} f) B_{q_0} f &+ \sum_{q_0+1}^{\infty} (f - \pi_q f) B_q f \\
&+ \sum_{q_0+1}^{\infty} (\pi_q f - \pi_{q-1} f) A_{q-1} f.
\end{aligned} \tag{4}
$$

The idea here is to write the left-hand side as the sum of $f - \pi_{q_1} f$ and $\sum_{q_0+1}^{q_1} (\pi_q f - \pi\theta_{q-1} f)$ for the largest $q_1 = q_1(f, x)$ such that each of the "links" $\pi_q f - \pi_{q-1} f$ in the "chain" is bounded in absolute value by $\sqrt{n}\, a_q$ (note that $|\pi_q f - \pi_{q-1} f| \leq \Delta_{q-1} f$). For a rigorous derivation note that either all $B_q f$ are zero or there is a unique q_1 with $B_{q_1} f = 1$. In the first case the first two terms in the decomposition are zero and the third term is an infinite series (all $A_q f$ equal 1) whose qth partial sum telescopes out to $\pi_q f - \pi_{q_0} f$ and converges to $f - \pi_{q_0} f$ by the definition of the $A_q f$. In the second case, $A_{q-1} f = 1$ if and only if $q \leq q_1$ and the decomposition is as mentioned, apart from the separate treatment of the case that $q_1 = q_0$, when already the first link fails the test.

Apply the empirical process $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ to each of the three terms in the decomposition (3) separately, take suprema over $f \in \mathscr{F}$ and use the triangle inequality to separate the three terms.

Since $|f - \pi_{q_0} f| B_{q_0} f \leq 2F 1\{2F > \sqrt{n}\, a_{q_0}\}$, the L_1-norm of the first supremum satisfies

$$
\begin{aligned}
\mathrm{E}^* \left\| \mathbb{G}_n (f - \pi_{q_0} f) B_{q_0} f \right\|_{\mathscr{F}} &\lesssim \sqrt{n}\, P F 1\left\{2F > \sqrt{n}\, a_{q_0}\right\} \\
&\lesssim a_{q_0}^{-1} P F^2 \lesssim 2^{-q_0} \sqrt{\log N_{q_0+1}}.
\end{aligned}
$$

Second, since the partitions are nested, $\Delta_q f B_q f \leq \Delta_{q-1} f B_q f$, which are bounded by $\sqrt{n}\, a_{q-1}$ for $q > q_0$. This implies

$$
P(\Delta_q f B_q f)^2 \leq \sqrt{n}\, a_{q-1} P \Delta_q f 1\left\{\Delta_q f > \sqrt{n}\, a_q\right\} \leq 2 \frac{a_{q-1}}{a_q} 2^{-2q}.
$$

Apply the triangle inequality and next the preceding lemma to find that

$$\begin{aligned}
&\mathrm{E}^*\left\| \sum_{q_0+1}^{\infty} \mathbb{G}_n(f-\pi_q f)B_q f\right\|_{\mathscr{F}} \\
&\quad \le \sum_{q_0+1}^{\infty} \mathrm{E}^*\|\mathbb{G}_n\Delta_q f B_q f\|_{\mathscr{F}} + \sum_{q_0+1}^{\infty} 2\sqrt{n}\,\|P\Delta_q f B_q f\|_{\mathscr{F}} \\
&\quad \lesssim \sum_{q_0+1}^{\infty}\left[a_{q-1}\log N_q + \sqrt{\frac{a_{q-1}}{a_q}}\,2^{-q}\sqrt{\log N_q} + \frac{2}{a_q}2^{-2q}\right].
\end{aligned}$$

Since a_q is decreasing, the quotient a_{q-1}/a_q can be replaced by its square. Then the series on the right-hand side can be bounded by a multiple of $\sum_{q_0+1}^{\infty} 2^{-q}\sqrt{\log N_q}$.

Third, there are at most $N_q - 1$ functions $\pi_q f - \pi_{q-1} f$ and at most $N_{q-1} - 1$ functions $A_{q-1}f$. Since the partitions are nested, the function $|\pi_q f - \pi_{q-1} f| A_{q-1} f$ is bounded by $\Delta_{q-1} f A_{q-1} f \le \sqrt{n}\, a_{q-1}$. Apply Lemma A.1 to the variables $\mathbb{G}_n(\pi_q f - \pi_{q-1} f)A_{q-1}f$ to find that

$$\mathrm{E}^*\left\| \sum_{q_0+1}^{\infty} \mathbb{G}_n(\pi_q f - \pi_{q-1} f)A_{q-1} f\right\|_{\mathscr{F}} \lesssim \sum_{q_0+1}^{\infty}\left[a_{q-1}\log N_q + 2^{-q-1}\sqrt{\log N_q}\right].$$

This is again bounded by a multiple of $\sum_{q_0+1}^{\infty} 2^{-q}\sqrt{\log N_q}$.

Finally, by similar arguments

$$\begin{aligned}
\mathrm{E}^*\|\mathbb{G}_n\pi_{q_0} f\|_{\mathscr{F}} &\lesssim E^*\left\|\mathbb{G}_n \pi_{q_0} f 1\left\{F \le \sqrt{n}\, a_{q_0}\right\}\right\|_{\mathscr{F}} + \sqrt{n}\, P^* F 1\left\{2F > \sqrt{n}\, a_{q_0}\right\} \\
&\lesssim a_{q_0}\log N_{q_0} + 2^{-q_0}\sqrt{\log N_{q_0}} + 2^{-q_0}\sqrt{\log N_{q_0+1}}.
\end{aligned}$$

Collecting the bounds of the last four paragraphs we see that the left-hand side of (1) can be bounded by a multiple of $\sum_{q_0+1}^{\infty} 2^{-q}\sqrt{\log N_q}$. In view of (3) this is bounded by a multiple of the entropy integral on the right of (1). □

Alexander (1987) first stated that the sum class of two Donsker classes $\mathscr{F}$ and $\mathscr{G}$, the class of all functions $x \to f(x) + g(x)$ when f and g range over $\mathscr{F}$ and $\mathscr{G}$, respectively, is Donsker. Since taking a sum is a Lipschitz operation on two arguments, this can be proved under measurability assumptions by extending Corollary 14.8 of Gine 3 and Zinn (1986a) to functions of two classes. However, the measurability assumptions are not needed. This was pointed out for uniformly bounded classes by Talagrand (1987). Here we shall give a short, direct proof, based on the fact that the convex hull of a Donsker class $\mathscr{F}$, the set of all functions $x \to \sum_{i=1}^{k} a_i f_i(x)$ with $k \in \mathbb{N}$, $\sum|a_i| \le 1$ and $f_i \in \mathscr{F}$, is Donsker by Dudley (1985).

THEOREM A.3. *The sum $\mathscr{F} + \mathscr{G}$ of two Donsker classes $\mathscr{F}$ and $\mathscr{G}$ is Donsker.*

PROOF. Since for arbitrary classes $\mathscr{F}_0 \subset \mathscr{F}$ the map ϕ: $l^\infty(\mathscr{F}) \to l^\infty(\mathscr{F}_0)$ defined by $\phi(z)(f_0) = z(f_0)$ is continuous, it is clear that a subset of a Donsker class is Donsker. Since the sum class is contained in a multiple of the convex hull of the union $\mathscr{F} \cup \mathscr{G}$ and a convex hull is Donsker by Dudley (1985), it suffices to show that $\mathscr{F} \cup \mathscr{G}$ is Donsker. We may without loss of generality assume that $\mathscr{F}$ and $\mathscr{G}$ are disjoint. Since $\mathscr{F}$ is Donsker there exists for every $\eta > 0$ a finite partition $\mathscr{F} = \bigcup_i \mathscr{F}_i$ such that

$$\limsup_{n \to \infty} \mathrm{E}^* \sup_i \sup_{f_1, f_2 \in \mathscr{F}_i} \left| \mathbb{G}_n(f_1 - f_2) \right| < \eta.$$

Similarly there is such a partition $\mathscr{G} = \bigcup_j \mathscr{G}_j$ of $\mathscr{G}$. Joining these partitions we obtain a partition of $\mathscr{F} \cup \mathscr{G}$ in finitely many sets $\mathscr{H}_i$ (with every $\mathscr{H}_i$ being either an $\mathscr{F}_i$ or a $\mathscr{G}_j$) with the property

$$\limsup_{n \to \infty} \mathrm{E}^* \sup_i \sup_{h_1, h_2 \in \mathscr{H}_i} \left| \mathbb{G}_n(h_1 - h_2) \right| < 2\eta.$$

This shows that the empirical process indexed by $\mathscr{F} \cup \mathscr{G}$ is asymptotically tight in $l^\infty(\mathscr{F} \cup \mathscr{G})$ and hence $\mathscr{F} \cup \mathscr{G}$ is Donsker [cf. Andersen and Dobrić (1987)]. □

Acknowledgments. I thank Richard Gill for posing the problem concerning monotone functions and the referees for a careful reading of the paper, as well as pointing out the connection to the Borisov–Durst theorem.

REFERENCES

ALEXANDER, K. S. (1987). The central limit theorem for empirical processes on Vapnik–Červonenkis classes. *Ann. Probab.* **15** 178–203.

ANDERSEN, N. T. and DOBRIĆ, V. (1987). The central limit theorem for stochastic processes. *Ann. Probab.* **15** 164–177.

ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542.

BIRMAN, M. S. and SOLOMJAK, A. (1967). Piecewise-polynomial approximations of functions of the classes W_p^α. *Math. USSR-Sb.* **73** 295–317.

DUDLEY, R. M. (1981). Donsker classes of functions. In *Proceedings of the Statistics and Related Topics Symposium* 341–352. North-Holland, Amsterdam.

DUDLEY, R. M. (1984). *A Course on Empirical Processes. Lecture Notes in Math.* **1097** 1–142. Springer, New York.

DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker classes, and weighted empirical processes. In *Probability in Banach Spaces V Lecture Notes in Math.* **1153** 141–178. Springer, New York.

DUDLEY, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probab.* **15** 1306–1326.

DUDLEY, R. M. (1990). Nonlinear functionals of empirical measures and the bootstrap. In *Probability in Banach Spaces VII* (E. Eberlein, J. Kuelbs and M. Marcus, eds.) 63–82. Birkhäuser, Boston.

DUDLEY, R. M. and KOLTCHINSKII, V. (1994). Envelope moment conditions and Donsker classes. Preprint.

GILL, R. D. (1994). Survival analysis. In *Ecole d'Eté de Probabilités de Saint-Flour XXII. Lecture Notes in Math.* **1581** 115–241. Springer, Berlin.

GINÉ, E. and ZINN, J. (1986a). Lectures on the central limit theorem for empirical processes. *Probability and Banach Spaces Lecture Notes in Math.* **1221** 50–113. Springer, New York.

GINÉ, E. and ZINN, J. (1986b). Empirical processes indexed by Lipschitz functions. *Ann. Probab.* **14** 1329–1338.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

KOLMOGOROV, A. N. and TIKHOMIROV, V. M. (1961). Epsilon-entropy and epsilon-capacity of sets in function spaces. *Amer. Math. Soc. Transl. Ser.* 2 **17** 277–364.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, Berlin.

OSSIANDER, M. (1985). A central limit theorem under metric entropy with L_2-bracketing. *Ann. Probab.* **15** 897–919.

POLLARD, D. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser. A* **33** 235–248.

POLLARD, D. (1989). A maximal inequality for sums of independent processes under a bracketing condition. Unpublished manuscript.

TALAGRAND, M. (1987). Measurability problems for empirical processes. *Ann. Probab.* **15** 204–221.

FACULTEIT WISKUNDE EN INFORMATICA
VRIJE UNIVERSITEIT
DE BOELELAAN 1081A
1081 HV AMSTERDAM
THE NETHERLANDS
E-MAIL: aad@cs.vu.nl